DOCUMENT RESUME

ED 127 334 TM 005 419

AUTHOR Vasta, Ross; Sarmiento, Robert F.
TITLE Liberal Grading Improves Evaluations But Not Performance.
PUB DATE [Apr 76]
NOTE 9p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Academic Achievement; Class Attendance; *College Students; *Comparative Analysis; Course Evaluation; Grades (Scholastic); *Grading; Higher Education; Student Evaluation; Student Evaluation of Teacher Performance; Study Habits; Teacher Evaluation

ABSTRACT

Two grading distributions--a liberal and a more stringent curve--were compared between two sections of an undergraduate psychology course. The two classes were team taught and treated as similarly as possible. Four unit exams and an optional final were administered, with one section consistently receiving a more lenient grade distribution. No differences in exam performance or self-reports of study behavior and attendance occurred between groups. Student evaluations of course and instructors were significantly higher in the class receiving more liberal grading. The relationship of the results to previous findings are discussed with regard to methodology and implications. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED127334

Liberal grading improves evaluations but not performance

Ross Vasta and Robert F. Sarmiento

State University of New York College at Brockport

U S DEPARTMENT OF HEALTH EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Paper presented at the meeting of the American Educational Research Association, San Francisco, April, 1976.

TM005 419

The use of norm-referenced or "curved" grade distributions is widespread in college courses--particularly, large ones. Although this form of grading has been found to be as effective as a criterion-referenced or absolute grading system in maintaining student performance (Williams, Pollack, & Ferguson, 1975), its parameters have not been investigated. In particular, the effects of the leniency of the grade distribution remain to be explored. For example, does more liberal grading affect the quality of student work and in which direction?

The relationship between grades and course/instructor evaluations by students is somewhat better understood. In general, most research has shown that such evaluations vary positively with student grades (e.g., Doyle & Whitely, 1974; Gessner, 1973; Kennedy, 1975). One difficulty with this research is that grades typically have been confounded with actual student performance. Whether it is the receipt of the grade, or the student's particular level of work or performance, which is responsible for the favorable evaluation has not been disentangled. One way to address this question is through the use of separate groups of students, whose performance does not differ, but who receive different grades. In contrast with a correlational approach (e.g., Kennedy, 1975), such an experimental analysis would permit conclusions relating the effects of grading on student evaluations.

Method

Subjects

Two sections of an undergraduate Developmental Psychology course were used, both meeting for 75 minutes on Tuesdays and Thursdays. One section,

meeting at 11:00 a.m., had 134 students, while the other, meeting at 2:00 p.m., had 116 students. Both sections met in the same room, covered the same material, at the same rate, presented by the same two instructors, and were otherwise treated as similarly as possible.

Procedure

The semester was divided into four units of equal length, each followed by a 40-item multiple-choice exam covering that unit, with an optional comprehensive final exam. Dr. Sarmiento taught the first and fourth units while Dr. Vasta was in charge of the second and third units. Classes consisted primarily of lectures, supplemented by films and slides, for both instructors. The students were informed that their final course grade would be the numerical average of the best four out of the five possible exams, with grade cutoffs based on the performance of the class. The independent variable of the experiment was the distribution of grades, with the morning class, selected randomly by Dr. Sarmiento, receiving a distribution more difficult than used in previous semesters of the course and the afternoon class receiving one more liberal. Dr. Vasta was unaware of which group got which distribution while he was teaching. For the first exam, the morning class had 8% A, 17% B, 38% C, 24% D and 13% E grades, while the afternoon class had 21% A, 27% B, 37% C, and 10% D and 5% E grades. Similar distributions were used for the second and third exams. However, at the end of the course, final grades were assigned according to the liberal distribution for both sections. Performance on the first exam represents a pretreatment comparison of the two groups since they had not been assigned any grades yet.

At each of the four unit exams, the students in both sections filled out a short, anonymous questionnaire concerning their studying behaviors and class attendance for that unit. During the fourth unit, they filled

out a course evaluation questionnaire and instructor evaluations for both professors, again anonymously.

The question of primary interest was -- What effect would grade distribution have on exam performance, studying behaviors and class attendance, and course/instructor evaluations?

Results

Exam Performance

Mean scores for the morning and afternoon sections on the first exam were 66.75% correct and 66.68% correct, respectively, with similar distributions. As mentioned previously, this constitutes a pretreatment comparison. Performance on the second, third and fourth exams was used to examine the effects of the grade distribution manipulation. On each of these exams, there were only slight differences between the groups in means and distributions and these were not statistically significant based on t-tests and Kolmogorov-Smirnov tests (Roscoe, 1969). More in depth analyses were also conducted.

The differential grading procedure resulted in some scores that would obtain the same letter grade in both sections, although high in the range of scores for that grade in one section and low in the other. On the assumption that a different letter grade would be more salient than differences within a grade category, these scores were removed and only those students in each section with scores that would give rise to different letter grades on the first exam were compared on performance for exam two. Again, no statistically significant differences in means or distributions were obtained. From this group, students who on exam 2 again had received the same scores but different letter grades were compared on exam 3, with no significant differences occuring. Of these, those with scores on exam 3 giving rise to different letter grades were

compared on exam 4 performance and again no significant differences were found. Thus, there is no evidence that the grading manipulation had any effect on exam performance.

Studying Behaviors and Attendance

The results of the questionnaires examining studying behaviors and class attendance showed no significant effects between groups. The percentage of students from each class returning the questionnaires remained approximately the same for all four unit exams.

Course/Instructor Evaluations

The instrument used to examine student evaluations of the course and instructors was the standard departmental questionnaire, not designed specifically for use in this study. Thus, some of the items were not relevant to the current investigation and were not used in the analysis. The proportions of students from each section returning the questionnaire were not significantly different.

Two of the items on the questionnaire bear directly on the effectiveness of the grade manipulation. One asked students to predict their final letter grade in the course and the other asked them to rate the course grading system from strict to lenient. Results for these items were consistent with the differences in grade distributions between the sections with the former item significantly different between groups based on a t-test.

Of the remaining relevant items, sixteen were evaluations of the course and sixteen of the instructor, repeated for the other instructor. There were also four demographic items assessing sex, grade point average, class and reason for taking the course.

Six of the 48 evaluative items showed significantly different mean

responses between the two sections, with $\alpha = 0.05$. On each of these items, the group receiving the more lenient distribution gave more favorable evaluations. Group differences on three more items approached significance and only on one of these, concerning the adequacy with which the exams covered the course material, did the group getting the difficult distribution give a more favorable rating.

The two instructors were evaluated differently by both groups but a t-test on difference scores between the two groups for the 16 items concerning the instructors showed no significant effect so these items were combined for further analysis. Considering only the direction of the differences between groups, not the magnitude of these differences, 24 of the 32 items concerning instructors showed a more favorable rating by the group receiving the lenient grade distribution, a result deviating significantly from chance expectations ($\chi^2 = 8.00$, 1 d.f., $p < .01$). Of the items concerning the course, 14 out of 16 were rated more favorably by the group with the lenient distribution, again a significant effect ($\chi^2 = 9.00$, 1 d.f., $p < .01$).

Discussion

The failure of the two grade distributions to differentially affect student exam performance or self-reports of actual study behaviors and attendance is intriguing. Since manipulation "checks" confirm that the desired aspects of the treatment were implemented, the absence of differential performance raises a number of questions. For example, the extent to which such distributions interact with individual student characteristics, or the extent to which these exams measure nonspecific academic skills, are issues for future research. Perhaps, it is most parsimonious to conclude

simply that the grade manipulation, although salient, was not effective in altering students' studying or exam behaviors, which have been developed and maintained over a considerable period of time. The potential effects of such grading practices over a longer period of time or courses remain to be investigated.

The positive effects of higher grades on student evaluations are congruent with earlier findings (e.g., Kennedy, 1975). Moreover, they suggest that grades, per se, can produce this relationship, independent of the students' actual performance.

This finding might be used to support the current trend toward liberal grading, since such a practice does not appear to adversely affect performance, yet increases evaluative responses. The generality of the current results, however, remains to be established with a variety of students, courses, and instructional methods. Perhaps, a better immediate implication of the present findings is the use of caution in interpreting student evaluations. Since such data are often used to assess teacher or method effectiveness, the identification of additional influencing variables question the wisdom of these practices.

References

Doyle, K.O. Jr., & Whitely, S.E. Student ratings as criteria for effective teaching. American Educational Research Journal, 1974, 11, 259-274.

Gessner, P.K. Evaluation of instruction. Science, 1973, 180, 566.

Kennedy, W.R. Grades expected and grades received--their relationship to student's evaluations of faculty performance. Journal of Educational Psychology, 1975, 67, 109-115.

Roscoe, J.T. Fundamental research statistics for the behavioral sciences. New York: Holt, Rinehart & Winston, Inc., 1969.

Williams, R.G., Pollack, M.J., & Ferguson, N.A. Differential effects of two grading systems on student performance. Journal of Educational Psychology, 1975, 67, 253-258